Volume 13 Number 12 1985 Nucleic Acids Research

Analysis of repetitive sequence elements containing tRNA-like sequences

Charles B.Lawrence, Donald P.McDonnell and W.Jay Ramsey

Department of Cell Biology, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

Received 30 April 1985; Accepted 29 May 1985

ABSTRACT

Several repetitive sequence elements from diverse species share extensive sequence homology with tRNA molecules. Analysis of the tRNA-like sequences within these elements suggest that they have originated from authentic tRNA sequences. Elements containing tRNA-like sequences can be divided into three distinct groups whose members share extensive sequence homology, have similar sequence organization and have unique species distribution. We suggest that these three groups represent independent examples of retroposon (4) families that have originated from tRNAs.

INTRODUCTION

Among the various dispersed repetitive sequence elements that occur in the genomes of eukaryotes is a class which appears to have originated from RNA transcripts and were subsequently inserted in the genome as DNA. These are exemplified by processed pseudogenes such as mouse α-globin (1), pseudogenes of small nuclear RNAs (2), "Alu" sequences which are processed pseudogenes of 7SL RNA (3), and retrovirus-like transposons which transpose through an RNA intermediate (18-19). Repetitive elements of this type have been called retroposons by Rogers (4,5) and retrotransposons by Boeke et al. (18) to reflect their RNA origin and apparent propensity to appear at many different locations in the genome of an organism through some mechanism of transposition.

Our group is interested in locating and characterizing sequences in hnRNA and mRNA molecules which are involved in regulating their expression. The suggestion by Milner et al. (6) that the presence of the brain identifier (ID) repetitive sequence element in hnRNA molecules may regulate their processing and expression prompted us to search the nucleic acid sequence data base to find gene sequences which contain copies of this element. Among the sequences which proved to have significant homology to the ID sequence were several tRNA sequences with greater than 80% homology. Rogers (5) has observed that sequences within the ID repetitive element and several other repeat families

can be folded into the clover-leaf secondary structure model characteristic of tRNA molecules.

In this paper, we survey a number of repetitive sequence elements for the presence of tRNA-like sequences. The results of this study suggest that tRNAs have given rise to different retroposon families on at least three independent occasions.

COMPUTER PROGRAMS

All computer analyses were performed on a WICAT 150 workstation running the HELIX Sequence Information System software (C. Lawrence, unpublished).

Data bank searches

Similarity searches were performed on the GenBank nucleic acid sequence data bank (release 21.0) using the program HOMOSRCH. This program compares

TABLE I
COMPILATION OF REPETITIVE SEQUENCE ELEMENTS

Code Name[a]	Description	Sequence[b] Start	Stop	Reference[c]
HUMALUB1	Human alu repeat	-	-	J00081
HUMALUB2	Human alu repeat	-	-	J00087
HUMPOMCR1-6	Human alu repeats	-	-	K00059-64
MUSRSBAM5	Mouse Bam5 repeat	-	-	J00063
MUSRSB2A	Mouse B2 repeat	180	410	K00131
MUSRSB2B	Mouse B2 repeat	80	310	K00132
MUSDIA	Mouse DI-2(20) repeat	-	-	(12)
MUSDIB	Mouse DI-2(34) repeat	-	-	(12)
MUSRSB1A	Mouse B1 repeat	-	-	J00628
MUSRSB1ALU(A)	Mouse B1 repeat	38	203	J00630
MUSRSB1ALU(B)	Mouse B2-like repeat	234	450	J00630
MUSRSB1ALU(C)	Mouse B2 repeat	506	700	J00630
RATGH1(A)	Rat B2 repeat in IVS2 of growth hormone gene	960	1160	J00739
RATGH1(B)	Rat B2 repeat in IVS2 of growth hormone gene	1130	1360	J00739
RATGH1(C)	Rat ID (brain identifier) sequence	1330	1460	J00739
RATRNAID1	Rat ID repeat	1090	1190	J01878
RATUG2A	Rat ID repeat near a U2 snRNA gene	880[d]	780[d]	K00034
RATTUBALPS	Rat ID repeat near tubulin pseudogene	1260	1360	J00799
RATCYC450	Rat ID repeat near cytochrome gene	470	550	J00718
RATPSBPA1	Rat ID repeat near steroid binding protein gene	140	220	V01257
HAMALU250	Hamster repeat	30	230	J00052
BOVPOMC1	Bovine A repeat near proopiomelanocortin gene	-	-	J00014
BOVPOMC2	Bovine B repeat near proopiomelanocortin gene	-	-	J00015
BOVPOMC4	Bovine C repeat near proopiomelanocortin gene	-	-	J00018
BOVPOMC6	Bovine D repeat near proopiomelanocortin gene	-	-	J00020
GTARPT	Goat A repeat in large IVS of beta-globin gene	-	-	(13)
GTCRPT	Goat C repeat in large IVS of beta-globin gene	-	-	(13)
GTCARPT	Goat C-A repeat in large IVS of beta-globin gene	-	-	(13)
CHKCR1RSA	Chicken CR1U1A repeat	-	-	J00841
XENOAX	Xenopus OAX repeat	-	-	(14)

[a] Code names are GenBank entry names for sequences with GenBank Accession numbers
[b] Positions of the sequence start and stop are relative to the first base in the GenBank sequence entry
[c] GenBank Accession numbers, except numbers in parathenses, are reference numbers
[d] Reverse-complement of GenBank sequence

all alignments between a query sequence and a data bank sequence and computes the best region of continuous homology (no deletions or insertions) within a given span length of nucleotides. The span used for all searches in this study is 41 nucleotides.

Dot matrix analysis

Dot matrix analyses were performed using the programs MTX (matrix calculation) and GMTX (graphical display). Homologies are scored in this program using a proportional matching method (7,8). For this study scores were calculated over a span of 25 nucleotides and homologies displayed which exceeded a minimum of between 55 to 60%.

Sequence alignments

Sequence alignments were calculated using a modification of the method described by Staden (8).

SEQUENCES USED IN THE STUDY

The repetitive sequence elements sequence used in this study are compiled in Table I. This list includes examples of human Alu elements, mouse B1 and B2 elements, rat ID elements and other repetitive elements from hamster, cow, goat, chicken and Xenopus.

A list of structural RNA sequences is compiled in Table II. This list includes the major small nuclear RNAs, small ribosomal RNAs and small cytoplasmic RNAs. For similarity searches of tRNAs, a compilation of 237 tRNA sequences from the RNA sequence file of the GenBank data bank release 21.0 was used.

TABLE II
COMPILATION OF RNA SEQUENCES

Code Name	Description	GenBank Accession #
RATUR1	Rat U1 small nuclear RNA	J01882
RATUR2	Rat U2 small nuclear RNA	K00781
RATUR3A	Rat U3A small nuclear RNA	J01884
RATUR3B	Rat U3B small nuclear RNA	K00780
RATUR4A	Rat U4 small nuclear RNA	K00782
RATUR5A	Rat U5A small nuclear RNA	K00783
RATUR6A	Rat U6 small nuclear RNA	K00784
RATURU4A	Rat U4A small nuclear RNA	K00477
RATURU4B	Rat U4B small nuclear RNA	K00478
RATURU4C	Rat U4C small nuclear RNA	K00479
HUMR7SL	Human 7SL RNA	V00588
RATRR58S	Rat 5.8S ribosomal RNA	J01881
RATRRA	Rat 5S ribosomal RNA	K01594
RATRR45S	Rat 4.5S RNA-I nuclear RNA	J01877
HUMROYA	Human HY1 gene (cytoplasmic Ro RNA)	V00584
HUMROYC	Human HY3 gene (cytoplasmic Ro RNA)	V00585

TABLE III
REPEATS WITH HOMOLOGY TO RNA SEQUENCES

Repeat[a]	RNA[b]	% Homology[c]
HUMALUBI	HUMR7SL	89
HUMALUB2	HUMR7SL	89
MUSRSB1A	HUMR7SL	77
HUMPOMCR1	HUMR7SL	92
HUMPOMCR2	HUMR7SL	87
HUMPOMCR3	HUMR7SL	94
HUMPOMCR4	HUMR7SL	86
HUMPOMCR5	HUMR7SL	89
HUMPOMCR6	HUMR7SL	88
HUMPOMCR1[d]	HUMROYC	68
MUSRSBIALU(A)	HUMR7SL	74
MUSDIA	RATRR45S	64
MUSDIB	HUMROYA	64
BOVPOMC2	RATRR45S	63
MUSRSB2A	HUMROYA	65
RATGH1(A)	HUMROYA	63

a Code name from Table I
b Code name from Table II
c Best continuous homology between two sequences over a span of 41 nucleotides
d Homology found with the reverse-complement of the sequence

SIMILARITY OF REPEAT SEQUENCE ELEMENTS TO RNA SEQUENCES

To identify RNA sequences which may have given rise to repetitive elements, we compared the sequences in a list of structural RNA sequences (Table II) for similarity to sequences in the list of repetitive sequence elements (Table I). For each comparison, the search algorithm reported the best region of continuous homology between the two sequences over a span of 41 nucleotides. Comparisons which yielded values of homology of 63% or greater over this span are reported in Table III. Several human sequences containing Alu repetitive elements and mouse sequences containing a B1 repetitive element had homology values of 74-90% with 7SL RNA as expected (3). Other elements had significant homology to 4.5SI nuclear RNA and human cytoplasmic RNAs. These homologies were examined further by dot matrix analysis which indicated that the significant homology extended over only a limited region of the RNA sequences (not shown). This suggests that the observed similarity is likely to be fortuitous and not indicative of a RNA origin for those repetitive elements. Thus, the search resulted only in finding similarity between Alu and B1 elements and the 7SL RNA which has been previously characterized.

The compilation of 237 tRNA sequences was then searched for similarity to the repetitive sequence elements. Sequences showing 68% or greater homology over a span of 41 nucleotides in each comparison are reported in Table IV. Only the best score for each repetitive sequence is shown. It should be noted

TABLE IV
REPEAT SEQUENCES SHOWING HOMOLOGY WITH tRNA

Repeat[a]	tRNAs[b]	% Homology[c]
MUSRSBAM5	RATMTTRLM	68
MUSRSB2A	RABTRK3	74
MUSRSB2B	LLUTRFN	73
MUSRSB1ALU(B)	DROTRE4	69
MUSRSB1ALU(C)	RABTRK3	76
RATGH1(A)	RABTRK3	73
RATGH1(B)	TACTRMM	75
RATGH1(C)	MZECPTRI2, SPNCPTRI2	80
RATTUBALPS	HCUTRA	79
RATUG2A	DROTRV4, ECOTRA1A, ECOTRA1B, HCUTRA	74
HAMALU25Q	RATTRS3	68
BOVPOMC1[d]	BSUTRK1, BSUTRK3	71
BOVPOMC2	BMOTRA1	71
BOVPOMC6	HUMTRGGCC	75
GTCRPT	HUMTRGGCC	72
GTCARPT	HUMTRGGCC	70
XENOAX	EGRCPTRF, RRUTRF	71

[a] Code name from Table I
[b] Code name from Table V
[c] Best continuous homology between two sequences over a span of 41 nucleotides
[d] Homology found with the reverse-complement of the sequence

that elements with significant homology to one tRNA also had signficant homology to several other tRNAs but with slightly lower scores. The best scores observed for the comparisons reached 80%. Table V contains a description of each tRNA identified in Table IV. To evaluate the significance of the scores, twelve random sequences 200 nucleotides in length were generated and searched against the list of tRNAs. A Z value (15-17) was then calculated for

TABLE V
tRNAS WITH HOMOLOGY TO REPEAT SEQUENCES

Code Name	Description	GenBank Accession #
BMOTRA1	Bombyx mori ala-tRNA-1	K00150
BSUTRK1,3	B. subtilis lys-tRNA-1 and -3	K00284,5
DROTRE4	D. melanogaster glu-tRNA-4	K00193
DROTRV4	D. melanogaster val-tRNA-4	K00250
ECOTRA1A,B	E. coli ala-tRNA-1A and B	K00139,40
EGRCPTRF	Euglena gracilis chloroplast phe-tRNA	K00340
HCUTRA	Halobacterium C. ala-tRNA	K00142
HUMTRGGCC	Human gly-tRNA-GCC	K00208
LLUTRFN	Lupin minor phe-tRNA	K00346
MZECPTRI2	Maize chloroplast ile-tRNA-2	K00223
RABTRK3	Rabbit lys-tRNA-3	K00291
RATMTTRLM	Rat mitochondrial Leu-tRNA	K00241
RATTRS3	Rat ser-tRNA-3	K00371
RRUTRF	R. rubrum phe-tRNA	K00331
SPNCPTRI2	Spinach chloroplast ile-tRNA-2	K00222
TACTRMM	T. acidophilum met-tRNA-M	K00302

TABLE VI
HOMOLOGY BETWEEN REPEATS WITH tRNA-LIKE SEQUENCES[a]

	Repeat	1[b]	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	GTCRPT	-	*[c]	*	*	*											
2	GTCARPT	84	-	*	*												
3	BOVPOMC2	76	79	-	*												
4	BOVPOMC6	78	84	88	-	*											
5	BOVPOMC1[d]	71	60	69	71	-											
6	MUSRSB2A	53	53	52	54	60	-	*	*	*	*	*					
7	MUSRSB2B	60	66	55	58	60	98	-	*	*	*	*					
8	MUSRSB1ALU(B)	65	67	60	60	57	72	91	-	*	*	*					
9	RATGH1(A)	52	56	52	54	58	97	98	76	-	*	*					
10	RATGH1(B)	54	56	56	52	58	98	98	76	100	-	*					
11	HAMALU250	60	60	62	61	55	95	97	88	98	98	-					
12	RATUG2A	62	51	63	63	52	57	59	50	61	57	59	-	*	*		
13	RATGH1(C)	57	57	60	60	50	62	64	56	65	65	63	94	-	*		
14	RATTUBALPS	58	61	64	61	52	63	61	58	63	68	62	87	98	-		
15	XENOAX	63	59	58	55	55	57	61	55	65	64	63	67	67	65	-	
16	MUSRSBAM5	59	49	52	50	43	58	57	55	54	55	57	51	51	52	54	-

[a] Values in table represent the best continuous homology between two sequences over a span of 41 nucleotides
[b] Numbers represent the number of the repetitive element in the first column
[c] Asterisks are placed for comparisons with a homology of 70% or greater
[d] Reverse-complement of sequence used

the homology values in Table IV where Z = (% homology - mean of % homology for random sequences)/(standard deviation of % homology for random sequences). The homologies in Table IV have Z values between 5.8 and 10.1 which are indicative of significant homology (15-17).

CHARACTERIZATION OF tRNA-LIKE SEQUENCES WITHIN REPETITIVE SEQUENCES

Repetitive sequence elements having homology to tRNA sequences were compared with each other to find related elements. The result of the comparison is compiled in Table VI. The XENOAX repeat and MUSRSBAM5 repeats do not show significant homology to other elements. The remainder of the elements, however, cluster into three groups: The goat and bovine repeats (numbers 1-5 in Table VI); the mouse, rat and hamster B2-like repeats (numbers 6-11); and the rat ID repeats (numbers 12-14). The members of each group have significant homology to each other but not to the other repetitive elements.

The tRNA-like sequences in representatives of each of the three groups defined above were examined by dot-matrix comparison of the repeat with a tRNA to which it shows significant homology. The resulting matrices are displayed in Figure 1 with only the tRNA-like region of the repeat shown. Mouse, rat and hamster representatives of the B2-like repeats were compared with rabbit lysine tRNA-3 (panels a, b and c); a rat ID sequence was compared with human glycine tRNA (panel d) and a bovine and goat repeat compared with a B. sub-

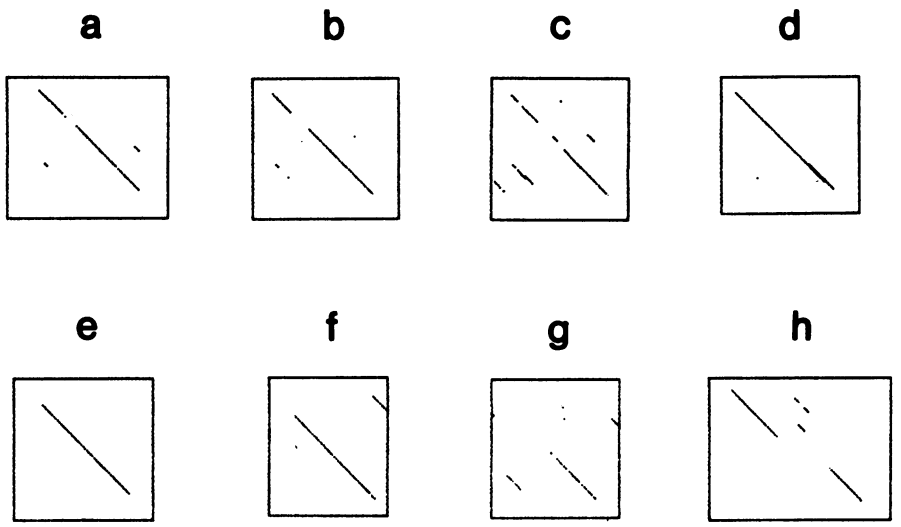


Figure 1

Dot matrix analysis of tRNA-like sequences in repetitive sequence elements. Dot matrix analysis between repetitive sequence elements and specific tRNA seuqences was performed. The region of the repetitive sequence is on the horizontal axis (h) and the tRNA is on the vertical axis(v). Panels: a. MUSRSB2A(h), RABTRK3(v); b. RATGH1(A)(h), RABTRK3(v); c. HAMALU250(h), RABTRK3(v); d. RATTUBALPS(h), HUMTRGGCC(v); e. BOVPOMC1(reverse-complement)(h), BSUTRK1(v); f. GTCARPT(h), HUMTRGGCC(v); g. MUSRSBAM5(h), RATMTTRLM(v); h. XENOAX(h), RRUTRF(v).

tilis lysine tRNA and human glycine tRNA respectively (panels e and f). All of these comparisons reveal homology of the repeat with most of the length of the tRNA sequence used in the comparison. By contrast, the MUSRSBAM5 repeat is homologous to only the 3'-half of rat mitochondrial leucine tRNA (panel

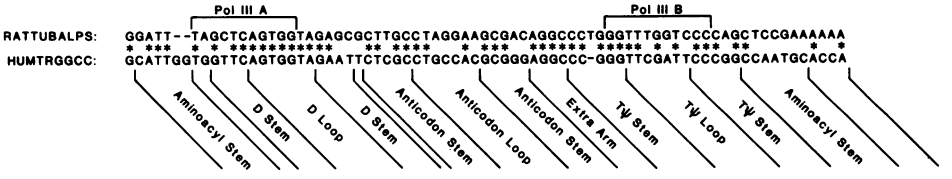
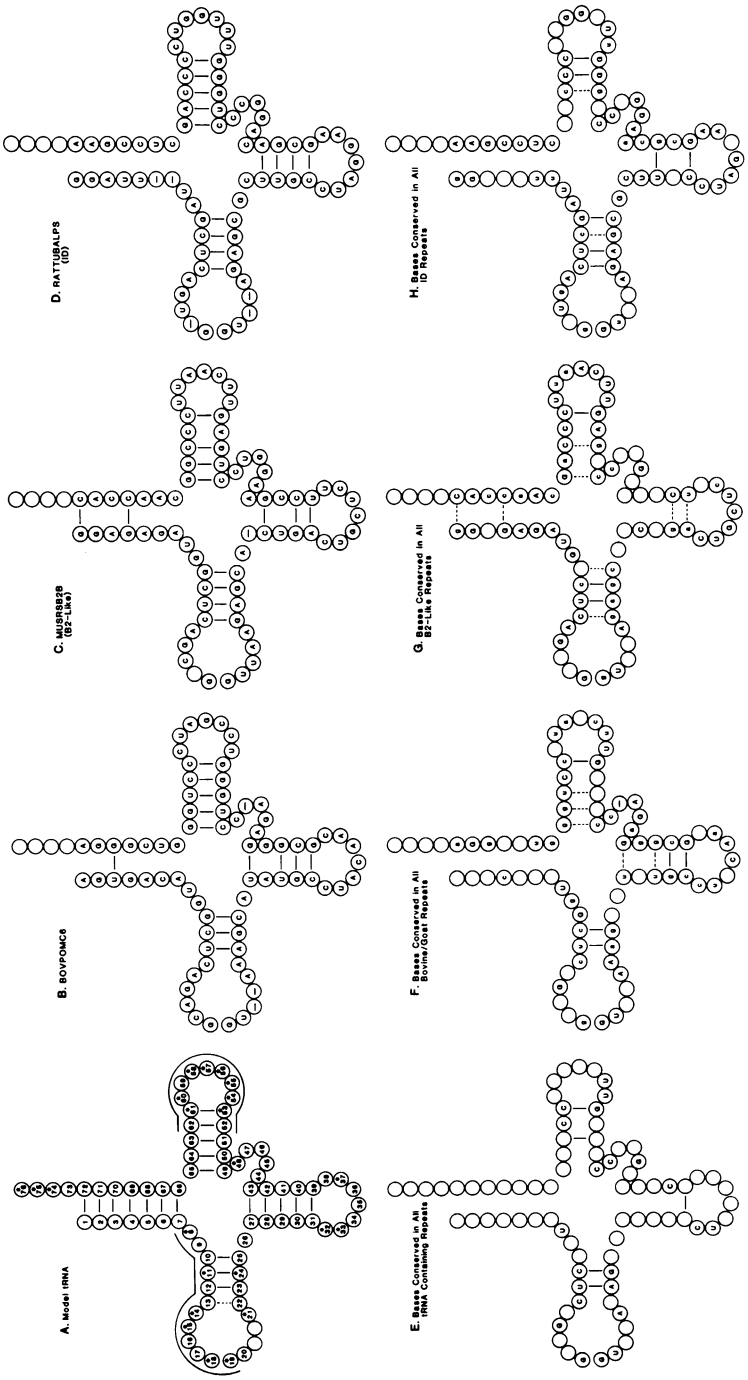


Figure 2

Alignment of tRNA-like sequence with a tRNA. The tRNA-like sequence in the ID repetitive element from the RATTUBALPS sequence was aligned with human glycine tRNA (HUMTRGGCC). The positions of the RNA polymerase III split promoter and landmarks in the tRNA sequence are indicated.


A. Model tRNA
B. BOVPOMC6
C. MUSHBB2B (B2-Like)
D. RATTUBALP6 (ID)
E. Bases Conserved in All tRNA Containing Repeats
F. Bases Conserved in All Bovine/Goat Repeats
G. Bases Conserved in All B2-Like Repeats
H. Bases Conserved in All ID Repeats

g). The Xenopus XENOAX repeat is homologous to the 5'- and 3'-thirds of the R. rubrum tRNA sequence with little homology to the central portion (panel h). Thus, dot-matrix analysis of representatives of the three groups defined in Table VI all show extensive homology to most of the length of specific tRNA molecules.

To examine the tRNA-like sequences within the three repeat groups in more detail we attempted to fold them into the clover-leaf secondary structure model common to all tRNAs. This was done by aligning the tRNA-like region of the repeat elements with a specific tRNA sequence to establish the homologous positions between the repeat sequence and the tRNA. An example of an alignment of a rat ID sequence with human glycine tRNA is shown in Figure 2. The nucleotides of the repeat sequence were then placed into the homologous positions of the clover-leaf model according to the assignments of Gauss and Sprinzl (9). Panels b, c and d of Figure 3 demonstrate this for examples of a bovine repeat, mouse B2-like repeat and rat ID repeat respectively. In all three cases potential base pairing in the D-stem, anti-codon stem and Tψ stem is preserved. This procedure was repeated for a total of five examples of the bovine/goat group, seven examples of the B2-like group and five examples of the rat ID group. The results are summarized in panels f, g and h of Figure 3. Positions having identical nucleotides in all examples are in upper-case; positions having identical nucleotides in all but one of the examples are in lower-case; potential base pairing present in all examples are indicated by a solid line and those present in all but one example are indicated by a broken line. The bovine and goat repeats maintain most of the base pairing potential in the anticodon and Tψ stems; B2-like sequences preserve base pairing in the

Figure 3
tRNA-like folding of repetitive sequence elements. The tRNA-like sequences in repetitive sequence elements were aligned with tRNA sequences to identify homologous positions between them. Using this alignment, the tRNA-like sequences were folded into the clover-leaf secondary structure model for tRNA according to the numbering system and alignments of Gauss and Sprinzl (9). Panel a: Numbering system for model tRNA. Asterisks represent invariant and semi-invariant bases in tRNAs, solid bars between nucleotides represent potential base pairing, solid bar on outsides of the D- and Tψ-stems and loops represent bases involved in the RNA polymerase III split promoter. Panels b-c: Folding of tRNA-like sequences in specific repetitive elements; BOVPOMC6 (panel b), MUSRSB2B (panel c), RATTUBALPS (panel d). Panels e-h: Conserved nucleotides and base pairing in groups of repetitive elements; all repetitive elements with tRNA-like sequences (panel e), bovine and goat elements (panel f), B2-like elements (panel g), ID elements (panel h). In panels f-h, nucleotides in upper-case letters or lower-case letters are present in all examples or in all but one example respectively; base pairing represented by solid bars or dashed lines are present in all examples or in all but one example respectively.

D stem; and ID sequences maintain the D stem and part of the Tψ stem. Panel e shows nucleotides and base-pairing preserved in all examples from all three groups. Base pairs between nucleotides 11-24, 12-23, 31-39, 51-63 and 53-61; and nucleotides at positions 8, 11, 12, 13, 15, 18, 19, 20, 21, 23, 24, 32, 33, 40, 45, 48, 49, 53, 54, 55, 61 and 62 are conserved in all three repetitive element groups. Of these conserved nucleotides, 14 are at positions which are invariant or semi-invariant in tRNA sequences (9).

To summarize this section, the tRNA-like sequences in the three repetitive sequence groups retain many of the potential base pairs characteristic of authentic tRNAs. This is not meant to imply that transcripts of the repetetive elements will adopt this secondary structure, but to point out remnants of the possible ancestors of the elements which have been conserved. In addition, nucleotides are conserved in all of these groups at 14 of the 20 invariant and semi-invariant positions (9) in tRNA. The region of highest similarity to tRNA is continuous from tRNA position 8 to 65 and includes the D, anticodon and Tψ stems and loops.

DOT MATRIX ANALYSIS OF REPETITIVE SEQUENCE ELEMENT GROUPS

The comparison of repetitive sequences that have regions homologous to tRNAs (Table VI) suggests the existence of three distinct groups of repeats that share extensive homology with other members of their group but not to members outside their group. We have characterized these repetitive element groups further with dot matrix analysis.

A sensitive method of defining the limits of a repetitive sequence element is to perform a dot matrix analysis on pairs of similar elements from different genetic origins. The homologous regions of the elements are visualized as a continuous diagonal in the matrix. The ends of the diagonal will define the ends of the region of homology and the limits of the repetitive sequence element.

Dot matrix analysis was performed on pairs of repetitive sequence elements from the three groups defined in Table VI. Typical results from some of these comparisons are shown in Figure 4. Panels a-e are examples of pairwise comparisons of members of the B2-like group. The region of homology between closely related mouse B2 elements (MUSRSBSA, MUSRSB2B) (panel a) extends for approximately 215 nucleotides. The MUSRSB2A B2 element is homologous to a rat B2 element [RATGH1(B)] over a similar region (panel b). Panels c and d show that a hamster B2 element (HAMALU250) is closely related to mouse and rat B2 elements except for a region of approximately 25 nucleotides in the 3'-half

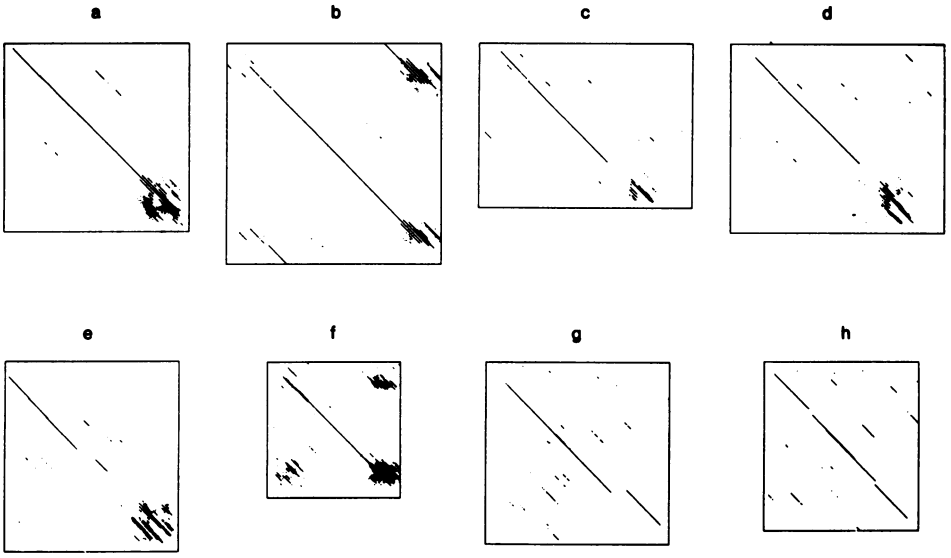


Figure 4
Dot matrix analysis of repetitive elements with other examples of the same type. Dot matrix analysis was performed to compare various repetitive sequence elements. The elements compared along with the positions of the first and last bases of each sequence displayed are as follows: panel a: MUSRSB2A(h)(180-410), MUSRSB2B(v)(80-310); panel b: MUSRB2A(h)(131-399), RATGH1(B)(v)(1106-1375); panel c: HAMALU250(h)(1-266), RATGH1(A)(v)(960-1160); panel d: HAMALU250(h)(1-266), MUSRSB2A(v)(180-410); panel e: MUSRSB1ALU(B)(h)(234-450), MUSRSB2A(v)(180-410); panel f: RATGH1(C)(h)(1314-1480), RATUG2A(v)(911-746); panel g: BOVPOMC2(h)(324-552), BOVPOMC6(v)(156-376); panel h: BOVPOMC6(h)(172-364), GTCARPT(v)(1-205).

of the element. Another example of a mouse repeat [MUSRSB1ALU(B)] is homologous to the 5'-half and 25 nucleotides at the 3'-terminus of the MUSRSBSA B2 repeat (panel e). The cluster of diagonals at the 3'-termini of these elements is due to the presence of an oligo(A) rich region which is characteristic of this repeat family (5). Panel f shows the comparison of an ID element from the second intervening sequence of rat growth hormone and one located in the flanking region of a U2 snRNA gene. The homologous region extends for approximately 130 nucleotides and terminates with an oligo(A) rich sequence (6). Panels g and h show the comparison of two bovine repeat elements and a bovine and a goat element respectively. These elements share a region of homology approximately 180 nucleotides in length in both bovine and goat examples. These elements do not terminate in an oligo(A) rich sequence. The common bovine/goat repeat unit is equivalent to the "C-A_3" element defined by Rogers (5).

DEFINITION OF THREE INDEPENDENT RETROPOSON FAMILIES ORIGINATING FROM tRNAS

Several criteria allowed us to define three independent families of retroposons containing tRNA-like sequences. Comparison of the sequence elements with themselves (Table VI) show that repetitive elements with tRNA-like sequences cluster into three groups whose members have homology with each other but do not have significant homology with members of the other groups. The three groups are: 1) a bovine and goat repetitive element group; 2) a group consisting of elements in the previously characterized B2 class; and 3) rat ID sequence elements.

Members of the bovine and goat group are approximately 180 bases long and have so far only been found in these two species. The tRNA-like portion of these elements is closely related to a human glycine tRNA. The 3'-terminus of the element is not rich in oligo(A).

Members of the B2-like group are approximately 215 nucleotides in length and are distrubuted in rodent species. All elements terminate in an oligo(A) rich sequence. The 5'-one-half to two-thirds portion of the elements are conserved with some sequence divergence observed in the 3'-portion of the element. The tRNA-like portion is closely related to a rabbit lysine tRNA.

Members of the ID group are found only in rats, are 130 nucleotides in length, and terminate in an oligo(A) rich sequence. The tRNA-like sequence in these elements is most closely related to a chloroplast isoleucine tRNA but is also very similar to human glycine tRNA.

All three groups contain sequences which appear to have been derived from tRNAs. The region of similarity with tRNA sequences is roughly the same for all three groups and includes the D, anticodon and Tψ stems and loops. They also retain many of the potential base pairs characteristic of tRNA molecules. Of particular note, is the observation that a large fraction of the invariant and semi-invariant positions in tRNAs are conserved in the homologous positions of the repetitive elements. Many of these lie outside of the split RNA polymerase III promoter. From these observations we consider it likely that the portion of the repetitive elements with homology to tRNA sequences is in fact derived from authentic tRNAs or their genes.

It is probably not possible to distinguish the particular tRNA which gave rise to the tRNA-like region of the repetitive elements because each repeat has significant homology to several closely related tRNA sequences. A small amount of sequence divergence from the original tRNA is likely to result in similarity to other tRNAs because of the similarity between tRNA sequences. The observation that the ID repeats are most similar to a chloroplast tRNA

probably does not reflect their origin, as they also are similar to several mammalian tRNAs, human glycine in particular.

The tRNA sequences for all three groups are located at the extreme 5'-boundary of the repetitive element and positioned such that transcription by RNA polymerase III, using the potential internal split promoter in the tRNA-like sequence, would initiate transcription close to the 5'-terminus of the element. This is strong evidence that the internal promoter in fact defines the 5'-boundary of the element and that transposition of the element occurs through an RNA intermediate. The potential internal promoters have been shown to be functional for the B2-like (10) and ID elements (11).

The structural similarity of members within each group of repetitive elements containing tRNA-like sequences and each group's unique species distribution and distinct structure argues that each of the three groups defined here has a distinct origin and evolutionary history. Each group contains the relic of a tRNA molecule within its sequence positioned in such a way that transcription by RNA polymerase III from the internal promoter in the tRNA-like sequence will result in a transcript which could serve as an intermediate in the (retro)transposition of the element to another genomic location. Possible mechanisms for this transposition have been reviewed (5). It appears that the generation of a retroposon family derived from a tRNA has occurred independently at least three times. It is likely that other examples of a similar event will eventually be observed.

ACKNOWLEDGEMENTS

The authors wish to thank Michelle Browner and Larry Rothblum for valuable comments, and Steve Delaune for typing the manuscript.

REFERENCES

1. Nishioka, Y., Leder, A. and Leder, P. Proc. Natl. Acad. Sci. USA. 77,2806-2809.
2. Van Arsdell, S.W., Denison, R.A., Bernstein, L.B., Weiner, A.M., Manser, T. and Gesteland, R.F. 1981. Cell 26,11-17.
3. Ullu, E. and Tschudi, C. 1984. Nature 312,171-172.
4. Rogers, J. 1983. Nature 301,460.
5. Rogers, J.H. 1984. Int. Rev. Cytol. [Suppl.] (in press).
6. Milner, R.J., Bloom, F.E., Lai, C., Lerner, R.A. and Sutcliffe, J.G. 1984. Proc. Natl. Acad. Sci. USA 81,713-717.
7. McLachlan, A.D. 1971. J. Mol. Biol. 61,609-424.
8. Staden, R. 1982. Nuc. Acids Res. 10,2951-2961.
9. Gauss, D.H. and Sprinzel, M. 1983. Nuc. Acids Res. 11,r1-r53.
10. Gutierrez-Hartmann, A., Leiberburg, I., Gardner, D., Baxter, J.D. and Cathala, G.G. 1984. Nucl. Acids Res. 12,7153-7173.

11. Sutcliffe, J.G., Milner, R.J., Gottesfeld, J.M. and Lerner, R.A. 1984. Nature 308,237-241.
12. Feagin, J.E., Setzer, D.R. and Schimke, R.T. 1983. J. Biol. Chem. 258,2480-2487.
13. Schon, E.A., Cleary, M.C., Haynes, J.R. and Lingrel, J.B. 1981. Cell 27,359-369.
14. Ackerman, E.J. 1983. EMBO Journal 2,1417-1422.
15. Doolittle, R. 1981. Science 214,149-159.
16. Wilbur, W.J. and Lipman, D.J. 1983. Proc. Natl. Acad. Sci. USA 80,726-730.
17. Lipman, D.J. and Pearson, W.R. 1985. Science 227,1435-1441.
18. Boeke, J.D., Garfinkel, D.J., Styles, C.A. and Fink, G.R. 1985. Cell 40,491-500.
19. Baltimore, D. 1985. Cell 40,481-482.